

Chris K. Ridder (SBN 218691)
Ridder, Costa & Johnstone LLP
chris@rcjlawgroup.com
440 N. Barranca Avenue, #7550
Covina, CA 91723
Tel: (650) 466-0586
Fax: 650) 466-6182

Rebecca Tushnet (No. 3043296 (N.Y.)) (pro hac vice
pending)
rtushnet@law.harvard.edu
520 Hauser, Harvard Law School
1575 Massachusetts Avenue
Cambridge, MA 02130
Telephone: 703-593-6759
Attorneys for Amici Curiae

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

RICHARD KADREY, *et al.*,
Individual and Representative Plaintiffs,

v.

META PLATFORMS, INC., a
Delaware corporation,

Defendant.

Case No. 3:23-cv-03417-VC-TSH

**[PROPOSED] AMICUS BRIEF OF
INTELLECTUAL PROPERTY
LAW PROFESSORS**

Date: May 1, 2025
Time: 10:00 a.m.
Dept: Courtroom 4, 17th Floor
Judge: Vince Chhabria
Trial Date: None
Date Action Filed: July 7, 2023

Table of Contents

TABLE OF AUTHORITIES.....	iii
STATEMENT OF INTEREST OF AMICI CURIAE	1
SUMMARY OF ARGUMENT.....	1
ARGUMENT	2
I. FAIR USE IS CONCERNED WITH THE PRODUCTION OF KNOWLEDGE, NOT “FAIRNESS” IN GENERAL.	2
II. COPYING TO CREATE NEW TOOLS IS TRANSFORMATIVE	5
III. GIVEN TRANSFORMATIVENESS, FAIR USE FACTORS TWO AND THREE ARE OF LIMITED SIGNIFICANCE.	7
IV. INTERMEDIATE COPYING FOR A TRANSFORMATIVE PURPOSE IS NOT A MARKET TO WHICH PLAINTIFFS ARE ENTITLED.	8
V. CONCLUSION	12
APPENDIX (SIGNATORIES).....	14

TABLE OF AUTHORITIES

Cases

<i>A.V. ex rel. Vanderhye v. iParadigms, LLC</i> , 562 F.3d 630 (4th Cir. 2009)	5, 6, 7
<i>American Institute of Physics v. Schwegman, Lundberg & Woessner, P.A.</i> , 2013 WL 4666330 (D. Minn. Aug. 30, 2013).....	7
<i>American Society for Testing & Materials v. Public.Resource.org, Inc.</i> , 82 F.4th 1262 (D.C. Cir. 2023).....	11
<i>American Society for Testing & Materials v. Public.Resource.org, Inc.</i> , 82 F.4th 1262 (D.C. Cir. 2023).....	11
<i>Arica Inst., Inc. v. Palmer</i> , 970 F.2d 1067 (2d Cir. 1992)	3
<i>Authors Guild v. Google, Inc.</i> , 770 F.Supp.2d 666 (S.D.N.Y.2011).....	14
<i>Authors Guild v. Google, Inc.</i> , 804 F.3d 202 (2d Cir. 2015).....	2, 10, 11
<i>Authors Guild, Inc. v. HathiTrust</i> , 755 F.3d 87 (2d Cir. 2014).....	6
<i>Bell v. Eagle Mountain Saginaw Independent School District</i> ; 27 F.4th 313 (5th Cir. 2022)	10
<i>Bill Graham Archives v. Dorling Kindersley Ltd.</i> , 448 F.3d 605 (2d Cir. 2006).....	10
<i>Blanch v. Koons</i> , 467 F.3d 244 (2d. Cir. 2006).....	8
<i>Cambridge University Press v. Patton</i> , 769 F.3d 1232 (11th Cir. 2014).....	11
<i>Campbell v. Acuff-Rose Music, Inc.</i> , 510 U.S. 569 (1994)	4
<i>Castle Rock Ent., Inc. v. Carol Pub. Grp., Inc.</i> , 150 F.3d 132 (2d Cir. 1998).....	2
<i>Flo & Eddie, Inc. v. Sirius XM Radio, Inc.</i> , 821 F.3d 265 (2d Cir. 2016)	5
<i>Google LLC v. Oracle America Inc.</i> , 593 U.S. 1 (2021).....	7
<i>Harper & Row Publishers, Inc. v. Nation Enterprises</i> , 471 U.S. 539 (1985).....	3
<i>Kelly v. Arriba Soft Corp.</i> , 336 F.3d 811 (9th Cir. 2003).....	9
<i>Lewis Galoob Toys, Inc. v. Nintendo of America, Inc.</i> , 780 F. Supp. 1283 (N.D. Cal. 1991) <i>aff'd</i> , 964 F.2d 965 (9th Cir. 1992)	9
<i>Perfect 10, Inc. v. Amazon.com, Inc.</i> , 508 F.3d 1146 (9th Cir. 2007).....	4, 11
<i>Sega Enterprises Ltd. v. Accolade Inc.</i> , 977 F.2d 1510 (9th Cir. 1992).....	7
<i>Swatch Grp. Mgmt. Servs. Ltd. v. Bloomberg L.P.</i> , 756 F.3d 73 (2d Cir. 2014).....	7
<i>Thomson Reuters Enterprise Centre GMBH v. Ross Intelligence Inc.</i> , --- F.Supp.3d ----, 2025 WL 458520 (D. Del. Feb. 11, 2025)	6

Walsh v. Townsquare Media, Inc., 464 F. Supp. 3d 570 (S.D.N.Y. 2020) 9

Statutes

U.S. Const., art. I, § 8, cl. 8 2

Other Authorities

Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem*, 93 Wash. L. Rev. 579 (2018)..... 11, 12

Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 Texas Law Review 743 (2021) 12

Matthew Sag, *Fairness and Fair Use in Generative AI*, 92 Fordham L. Rev. 1887 (2024) 4

Matthew Sag, *Copyright and Copy-Reliant Technology*, 103 N.W.U. L. REV. 1607 (2009)10

Michael Masnick & Leigh Beadon, *The Sky Is Rising 2024: A Detailed Look at the State of the Entertainment Industries* 12

Pamela Samuelson, *Fair Use Defenses in Disruptive Technology Cases*, 71 UCLA L. Rev. 1484 (2024) 11

Pierre N. Leval, *Toward a Fair Use Standard*, 103 Harv. L. Rev. 1105 (1990) 4

Rebecca Tushnet, *All of This Has Happened Before and All of This Will Happen Again: Innovation in Copyright Licensing*, 29 Berkeley Tech. L.J. 1447, 1483 (2014)..... 11

Molly S. Van Houweling, *The Freedom to Extract in Copyright Law*, 103 N.C. L. REV. 445 (2025).....9-10

Amici have sharply different views on the social costs and benefits of generative AI, but agree on one thing: Copyright is not the correct tool for managing those costs and benefits, especially with respect to training data. Rejecting existing precedent about the fairness of intermediate uses—internal copying designed to produce a noninfringing output—would be harmful to the copyright system as a whole, which is designed to encourage the production of new knowledge and new works.

STATEMENT OF INTEREST OF AMICI CURIAE

Amici, listed in the Appendix are scholars whose research and teaching focus has included copyright law.¹ Amici have no direct interest in the outcome of this litigation.

SUMMARY OF ARGUMENT

Plaintiffs’ motion for partial summary judgment argues that the fair use analysis related to Meta’s internal copying should be analyzed in isolation from the purpose of that internal reproduction: to create a tool that will create new outputs. This is a legal error; the purpose of the copying can only be analyzed along with its purpose: to create a new tool that generates new insights and information.²

¹ Amici certify that no party’s counsel authored this brief in whole or in part; no party or party’s counsel contributed money that was intended to fund the preparation or submission of this brief; and no person other than the amici contributed money that was intended to fund the preparation or submission of this brief.

² Amici take no position on whether a specific output infringes or whether a fair use defense is available to specific outputs. Amici also are not arguing that Meta may share full or substantially similar copies with the public.

The case law, including binding circuit precedent, holds that internal copying, made in the course of creating new knowledge, is a transformative use that is heavily favored by fair use doctrine. In addition, the cases reject any claimed market for licensing non-consumptive, non-expressive use; the desire to be paid for use does not create a market. These are the key issues before the Court.

ARGUMENT

I. Fair Use Is Concerned with the Production of Knowledge, Not “Fairness” in General.

“[W]hile authors are undoubtedly important intended beneficiaries of copyright, the ultimate, primary intended beneficiary is the public, whose access to knowledge copyright seeks to advance by providing rewards for authorship.” *Authors Guild v. Google, Inc.*, 804 F.3d 202, 212 (2d Cir. 2015) [*Authors Guild*]. Thus, “[t]he ultimate test of fair use is whether the copyright law’s goal of ‘promoting the Progress of Science and useful Arts’ ‘would be better served by allowing the use than by preventing it.’ ” *Castle Rock Ent., Inc. v. Carol Pub. Grp., Inc.*, 150 F.3d 132, 141 (2d Cir. 1998) (quoting U.S. Const., art. I, § 8, cl. 8; *Arca Inst., Inc. v. Palmer*, 970 F.2d 1067, 1077 (2d Cir. 1992)).

That is, copyright, including fair use, does not seek to benefit the public by improving life in general: it seeks to benefit the public *by increasing access to knowledge*. The Supreme Court has endorsed this view: The “public benefits the copying will likely produce” are the focus of the assessment, as they are “related to copyright’s concern for the creative production of new expression.” *Google*, 593 U.S. at 35.

Modern copyright cases have, in line with this reasoning, considered the defendant’s “good faith” at most when it related to intentional displacement of the right of first publication. In *Harper & Row, Publishers, Inc. v. Nation Enterprises*, the Supreme Court held that the Nation’s intended purpose to “scoop” other publishers and thus supplant the right of first publication was relevant to the first fair use factor analysis. 471 U.S. 539, 562 (1985). More recent copyright cases have expressed skepticism about the nebulousness of “good faith” outside the more manageable question of whether there is substitution for a legitimate, functioning licensing market for a specific use. *Google*, 593 U.S. at 32; *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 585 n.18 (1994).³ Thus, it is wrong to say, as plaintiffs do, that “for fair use to apply, the work that was copied must have been lawfully acquired in the first place.” Kadrey Br. at 22.

Plaintiffs’ argument conflates the work (the subject of copyright) with the unauthorized making of a specific copy of the work (a precondition for *any* claim that the right of reproduction was infringed). Just as possessing a lawfully made copy does not itself entitle a copy-owner to make more copies, possessing a lawfully made copy is not a precondition for fair use. In *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146 (9th Cir. 2007), for example, the Ninth Circuit found fair use even though Google’s thumbnails were produced by copying unauthorized copies on “pirate” sites and linked back to those sites. *Id.* at 1157, 1166. Nonetheless, the *product* of Google’s copying—the thumbnail

³ See also, e.g., *Swatch Group Management Services Ltd. v. Bloomberg L.P.*, 756 F.3d 73, 82-84 (2d Cir. 2014) (unauthorized access did not weigh against fair use where use was transformative); *Santos v. Kimmel*, 745 F. Supp. 3d 153, 165 (S.D.N.Y. Aug. 19, 2024) (deception and contract violation didn’t weigh against fair use).

1 images and image search database—was transformative and therefore the copying was
2 fair. *Id.* at 1165-66.⁴

3 The cases cited by plaintiffs, by contrast to the case at bar, involve unauthorized
4 copies made for consumptive use—downloads that were then enjoyed as individual
5 works by the downloaders. Kadrey Br. 23-24, 26-27. Those cases do not set forth a rule
6 for copies that were never perceived by humans, which were made in order to make new
7 tools and identify new insights. The cases instead agree that such “big data” copies are
8 transformative. *See* Matthew Sag, *Fairness and Fair Use in Generative AI*, 92 Fordham
9 L. Rev. 1887, 1905 (2024) (explaining how the case law distinguishes between uses that
10 systematically expose substantial amounts of expression to the public and uses that do
11 not); *cf. Flo & Eddie, Inc. v. Sirius XM Radio, Inc.*, 821 F.3d 265, 270 n.4 (2d Cir. 2016)
12 (“The fair-use analysis applicable to [creating an internal database of recordings] ... is
13 bound up with whether the ultimate use of the internal copies [to make public
14 performances] is permissible.”).

15 Fair use, like copyright as a whole, “is not a privilege reserved for the well-
16 behaved.” Pierre N. Leval, *Toward a Fair Use Standard*, 103 Harv. L. Rev. 1105, 1126
17 (1990). Fair use doctrine should focus on the consequences of a ruling for knowledge and
18 expression. Other considerations should be left for other legal regimes.
19
20
21
22
23
24

25 ⁴ *See also id.* at 1164 n.8 (“We reject at the outset Perfect 10’s argument that providing
26 access to infringing websites cannot be deemed transformative and is inherently not fair
27 use.”).

II. Copying to Create New Tools Is Transformative

Matthew Sag has succinctly summarized the state of the case law:

Courts have consistently held that technical acts of copying that do not communicate an author's original expression to a new audience constitute fair use. Examples of non-expressive uses include copying object code to extract uncopyrightable facts and interoperability keys ("reverse engineering"), an automated process of copying student term papers to compare to other papers for plagiarism detection, copying HTML webpages to make a search engine index, copying printed library books to allow researchers to conduct statistical analyses of the contents of whole collections of books, and copying printed library books to create a search engine index.

The case law indicates that even though these non-expressive uses involved significant amounts of copying, they did not interfere with the original expression that copyright is designed to protect. Each use involved copying as an intermediate step toward producing something that either did not contain the original expression of the underlying work or contained a trivial amount. Courts have consistently held that non-expressive uses (although not labeled as such) are fair use.

Sag, *supra*, at 1903 (footnotes omitted); *see, e.g., Authors Guild*, 804 F.3d at 207, 214, 216-17 (enabling text analysis of a corpus of works by users was transformative; intermediate copying in that case allowed Google to "augment public knowledge ... without providing the public with a substantial substitute for matter protected by the Plaintiffs; copyright interests"); *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 90 (2d Cir. 2014) [*HathiTrust*] (same); *A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630, 639-40 (4th Cir. 2009) (creating plagiarism detection tool for use by schools and teachers was transformative); *see also Sega Enterprises Ltd. v. Accolade Inc.*, 977 F.2d 1510 (9th Cir. 1992) (internal copying of video games for analysis and creation of a noninfringing alternative); *American Institute of Physics v. Schwegman, Lundberg & Woessner, P.A.*, 2013 WL 4666330, *6-7, *10-11 (D. Minn. Aug. 30, 2013) (internal copying for the purpose of preparing patent applications was "transformative use because

it was related to patenting, not to using the expression of the articles”).⁵ These cases demonstrate that the intermediate use/non-expressive use cases are not limited to software, contrary to the suggestion of the court in *Thomson Reuters Enterprise Centre GMBH v. Ross Intelligence Inc.*, --- F.Supp.3d ----, 2025 WL 458520 (D. Del. Feb. 11, 2025), *motion for interlocutory appeal filed*, Mar. 18, 2025. There is no logical reason for any such limitation, and Professor Sag’s analysis explains why the principle is a general one.

Relatedly, the Supreme Court held that computer code that, though it actually directly copied code, also opened up significant new opportunities for people using that code, was also transformative. *Google LLC v. Oracle America Inc.*, 593 U.S. 1, 30 (2021) (“[Google’s] new product offers programmers a highly creative and innovative tool [I]ts use was consistent with that creative ‘progress’ that is the basic constitutional objective of copyright itself.”). The Court made clear that Java and Android had some overlapping uses and functions, but because Android was also generative of new programs and insights, it was transformative. *Id.* Even imperfect new tools are fair if their *purpose* is transformative. *A.V.* likewise held that summary judgment was appropriate despite a dispute over whether defendant’s plagiarism detection software actually worked. “The question of whether a use is transformative does not rise or fall on whether the use perfectly achieves its intended purpose.” 562 F.3d at 640.

⁵ Where the defendant’s internal copying was in the service of creating a work for distribution to the public, courts also hold that only the final version of the defendant’s work, and not intermediate versions, are the appropriate subject of analysis, even outside the fair use context. *See, e.g., Briggs v. Blomkamp*, 70 F. Supp. 3d 1155, 1168 n.2 (N.D. Cal. 2014), *aff’d sub nom. Briggs v. Sony Pictures Entm’t, Inc.*, 714 F. App’x 712 (9th Cir. 2018).

III. Given Transformativeness, Fair Use Factors Two and Three Are of Limited Significance.

Factor two, the nature of the work, is of less importance when the copying is internal and the public is not invited to consume works for their expressive value. *See, e.g., A.V.*, 562 F.3d at 641-42 (affirming conclusion that “even if the plaintiffs’ works were highly creative in nature, iParadigms’ use of the plaintiffs’ works was not related to the creative core of the works,” so factor two did not disfavor fair use). The degree of creativity in a work is also less important when a work has been previously published, so that the copyright owner has already had the opportunity to exploit it. *See, e.g., Swatch Grp. Mgmt. Servs. Ltd.*, 756 F.3d at 89 (“[B]ecause Swatch Group publicly disseminated the spoken performance embodied in the recording before Bloomberg’s use, the publication status of the work favors fair use.”); *Blanch v. Koons*, 467 F.3d 244, 256 (2d Cir. 2006); *Kelly*, 336 F.3d at 820 (“The fact that a work is published or unpublished also is a critical element of its nature. Published works are more likely to qualify as fair use because the first appearance of the artist’s expression has already occurred.”) (footnote omitted); *Arica Institute, Inc. v. Palmer*, 970 F.2d 1067, 1078 (2d Cir. 1992) (factor two favors fair use where accusing work is “a published work available to the general public”); *Walsh v. Townsquare Media, Inc.*, 464 F. Supp. 3d 570, 585 (S.D.N.Y. 2020) (use of a previously published work with mixed creative and factual elements favors fair use); *Lewis Galoob Toys, Inc. v. Nintendo of America, Inc.*, 780 F. Supp. 1283, 1293 (N.D. Cal. 1991) (“The works’ published nature supports the fairness of the use.”), *aff’d*, 964 F.2d 965 (9th Cir. 1992).

Likewise, when internal, intermediate copying occurs for the transformative purpose of creating new knowledge, courts have routinely found fair use when an entire

work is copied. *See, e.g., Perfect 10, Inc.*, 508 F.3d at 1167–68 (use of entire image was necessary since using less would diminish usefulness of visual search engine which had transformative purpose) (citing *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 821 (9th Cir. 2003)); *Sag, supra*, at 1915 (“Although non-expressive uses typically involve making complete literal copies, courts have found such copying is reasonable when it is an intermediate technical step in an analytical process that does not lead to the communication of the underlying original expression to a new audience.”) (footnote omitted).

This tolerance for internal copying makes sense given the nature of the fair use inquiry. The core question is not whether, in judicial hindsight, the defendant took more than absolutely necessary, but whether the amount taken was “reasonable in relation to the purpose of the copying.” *Campbell*, 510 U.S. at 586; *see also Oracle*, 141 S.Ct. at 1205 (“Google copied those lines not because of their creativity, their beauty, or even (in a sense) because of their purpose. ... The ‘substantiality’ factor will generally weigh in favor of fair use where, as here, the amount of copying was tethered to a valid, and transformative, purpose.”).

IV. Intermediate Copying for a Transformative Purpose Is Not a Market To Which Plaintiffs Are Entitled.

Courts have repeatedly held that the fourth fair use factor does not disfavor a defendant when a defendant makes full copies of works, as long as the use was sufficiently transformative. *See, e.g., Authors Guild*, 804 F.3d at 223-25; *HathiTrust*, 755 F.3d at 98 (2nd Cir. 2014); *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 613 (2d Cir. 2006); *see also, e.g., American Society for Testing & Materials v.*

1 *Public.Resource.org, Inc.*, 82 F.4th 1262, 1271 (D.C. Cir. 2023); *A.V. v. iParadigms,*
 2 *LLC*, 562 F.3d 630, 642 (4th Cir. 2009); *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d
 3 1146, 1165 (9th Cir. 2007).

4 Courts’ consistent recognition that not all claimed losses are cognizable under
 5 factor four prevents the market harm factor from becoming circular. *Dorling Kindersley,*
 6 448 F.3d at 614-15 (“a copyright holder cannot prevent others from entering fair use
 7 markets”); *Castle Rock Entertainment Inc. v. Carol Publishing Group*, 150 F.3d 132, 146
 8 n.11 (2d Cir. 1998) (holding that a copyright owner cannot control fair use markets
 9 merely “by developing or licensing a market for parody, news reporting, educational or
 10 other transformative uses of its own creative work”); *Cambridge University Press v.*
 11 *Patton*, 769 F.3d 1232, 1276 (11th Cir. 2014) (“The goal of copyright is to stimulate the
 12 creation of new works, not to furnish copyright holders with control over all markets.
 13 Accordingly, the ability to license does not demand a finding against fair use.”); *id.* at
 14 1278 (“Plaintiffs may not head off a defense of fair use by complaining that every
 15 potential licensing opportunity represents a potential market for purposes of the fourth
 16 fair use factor.”).⁶

17
 18 In particular, the market for authorizing non-expressive internal uses is not within
 19 the scope of copyright owners’ rights, and so there is no cognizable harm when they are
 20
 21

22
 23 ⁶ Relatedly, the market harm must be material or substantial before it weighs against fair
 24 use. *Patton*, 769 F.3d 1282 (factor four “asks whether the market harm caused by
 25 Defendants’ unpaid copying will materially impair Plaintiffs’ incentive to publish”;
 26 *Authors Guild v. Google, Inc.*, 804 F.3d at 224 (“[T]he possibility, or even the probability
 27 or certainty, of some loss of sales does not suffice to make the copy an effectively
 28 competing substitute that would tilt the weighty fourth factor in favor of the rights holder
 in the original. There must be a meaningful or significant effect ‘upon the potential market
 for or value of the copyrighted work.’”) (quoting 17 U.S.C. § 107).

1 deprived of exclusivity over those uses. Although AI tools may create new works, if the
2 new works themselves are noninfringing, that is a *good* result from copyright's
3 perspective.

4 In order to appropriately protect copyrightable expression but not unprotected
5 elements within a work, factor four has not considered "market impacts that result from
6 the use of unprotected rather than protected elements." Molly S. Van Houweling, *The*
7 *Freedom to Extract in Copyright Law*, 103 N.C. L. Rev. 445, 509 (2025). A summary of
8 unprotected facts and ideas from a larger book may provide readers with all the
9 information they seek; a new Yellow Pages directory or other taxonomy may organize
10 unprotected data so well that previous versions are no longer attractive to readers. But
11 copyright does not protect against that kind of market loss. *NXIVM Corp.*, 364 F.3d at
12 482 ("[T]he relevant market effect with which we are concerned is the market for
13 plaintiffs' 'expression,' and thus it is the effect of defendants' use of that expression on
14 plaintiffs' market that matters, not the effect of defendants' work as a whole."); *Authors*
15 *Guild v. Google, Inc.*, 804 F.3d at 224 (cognizable market effect must be based on
16 copyrightable aspects of what was copied, not on uncopyrightable aspects). Recognizing
17 fair use in such cases is one way of implementing the Supreme Court's command that
18 factor four analysis must "take into account the public benefits the copying will likely
19 produce," particularly those "related to copyright's concern for the creative production of
20 new expression." *Oracle*, 593 U.S. at 35-36.

23 As Matthew Sag has explained:

24 [A]lthough the fourth factor risks collapsing into circularity because everything is
25 a potential market effect, courts have in fact avoided this nadir by applying certain
26 limiting principles that emphasize that the copyright market is limited to
27 expressive substitution. The logical implication of the exclusion of economic

consequences that do not arise from expressive substitution is that to the extent that a use is nonexpressive, it typically has no cognizable market effect under the fourth factor.... [F]air use cases often turn on the simple question of whether the particular market claimed by the plaintiff is one that is cognizable under copyright. ... This principle is reflected in the seemingly unrelated cases involving parody and the reverse engineering of computer software. In both scenarios, courts exclude consideration of market effects that do not arise from expressive substitution.

Copyright and Copy-Reliant Technology, 103 N.W.U. L. REV. 1607, 1653-54 (2009); *cf. Patton*, 769 F.3d at 1277-78 (risk of circularity is particularly high when the claimed market is “a market for licenses to use Plaintiffs’ works in a particular way”).

Large-scale intermediate copying is particularly unlikely to interfere with a cognizable market because of the sheer size of the corpus, which makes licensing unworkable as a practical matter and thus not traditional, reasonable, or likely to be developed. Pamela Samuelson summarized the problem:

[N]o class yet exists that is capable of granting a license to use class members’ works as training data, and none certainly existed when the defendants’ datasets were created and used to train AI models. Nor is it possible for AI developers to license the rights to use all in-copyright works available on the internet, given the exceptionally large number of works and copyright owners at stake. Transaction costs would be prohibitive relative to the value of use of works as training data. Hence, market failure considerations would seem to undercut the class action plaintiffs’ market harm arguments.

Pamela Samuelson, *Fair Use Defenses in Disruptive Technology Cases*, 71 UCLA L. Rev. 1484, 1561 (2024) (footnotes omitted). *See also* Rebecca Tushnet, *All of This Has Happened Before and All of This Will Happen Again: Innovation in Copyright Licensing*, 29 Berkeley Tech. L.J. 1447, 1483 (2014) (“To claim that licenses can replace fair use because *some* participants within each market are willing to license *most* of the time is to advocate the suppression of all fair uses that rely on works that aren’t within the licensing

scheme.”).⁷ Indeed, even an attempt to create a license-by-class-action regime for Google Books failed because of, inter alia, the substantial differences in authors’ interests that precluded class treatment. *Authors Guild v. Google, Inc.*, 770 F.Supp.2d 666, 679–680 (S.D.N.Y.2011).

These market failures have real consequences for the public benefits that large datasets can otherwise provide. Licensed alternatives are likely to have systematic biases in content that could do real harm. See Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem*, 93 Wash. L. Rev. 579, 590, 630 (2018) (explaining how limited, licensed datasets are likely to be biased against specific groups and perspectives); Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 Texas Law Review 743, 770 (2021) (“Smaller, proprietary datasets—particularly those with large and nonrandom gaps due to failures of copyright licensing—will lead to worse decisions by ML systems. And those worse decisions have real-world consequences.”) (footnote omitted).

V. Conclusion

Copyright owners have often predicted that new technologies, from photocopying to home VCRs to the internet, would create disasters for copyright owners and that fair use needed to be shrunk to protect them; instead, new technologies have routinely created

⁷ “In the music businesses, the one sector of copyrighted content headed to this model [of identifying and licensing everything], they are far from perfecting it despite nearly a century of good work towards it.” Nat’l Telecomm. & Info. Admin., Request For Comments On Department Of Commerce Green Paper, Copyright Policy, Creativity, And Innovation In The Digital Economy, No. 130927852-3852-01, Comments Of DeviantArt 31 (Nov. 13, 2013), available at http://www.ntia.doc.gov/files/ntia/deviant_art_comments.pdf.

new markets. *See* Samuelson, *supra*, at 1571-72; Michael Masnick & Leigh Beadon, The Sky Is Rising 2024: A Detailed Look at the State of the Entertainment Industries, <https://copia.is/library/the-sky-is-rising-2024/>.

This history should caution against rejecting the many precedents supporting intermediate copying for the purpose of creating new and useful tools that millions of people use. Whatever the risks of AI—and there may be many—condemning the act of creating large-scale training datasets as copyright infringement is not the answer.

Dated: March 31, 2025

By: /s/ Rebecca Tushnet
 Rebecca Tushnet (No. 3043296 (N.Y.))
 (pro hac vice pending)
 rtushnet@law.harvard.edu
 520 Hauser, Harvard Law School
 1575 Massachusetts Avenue
 Cambridge, MA 02130
 Tel: 703-593-6759

Chris K. Ridder (SBN 218691)
 Ridder, Costa & Johnstone LLP
 chris@rcjlawgroup.com
 440 N. Barranca Avenue, #7550
 Covina, CA 91723
 Tel: (650) 466-0586
 Fax: (650) 466-6182

Attorneys for Amici Curiae

APPENDIX⁸

Matthew Sag
Jonas Robitscher Professor of Law in Artificial Intelligence, Machine Learning, and Data
Science, Emory University.

Zahr K. Said
Professor of Law
Santa Clara University School of Law

Jessica Silbey
Professor of Law, Boston University School of Law

Rebecca Tushnet
Frank Stanton Professor of the First Amendment, Harvard Law School

⁸ Institutional affiliations are provided solely for purposes of identification.